基于双目视觉和改进YOLOv8的玉米茎秆宽度原位识别方法

左昊轩[1]，黄祺成[1]，杨佳昊[2]，孟繁佳[2]，李思恩[3]，李 莉[1*]

（1. 中国农业大学农业农村部农业信息获取技术重点实验室，北京 100083，中国；2. 中国农业大学智慧农业系统集成研究教育部重点实验室，北京 100083，中国；3. 中国农业大学 水利与土木工程学院，北京 100083，中国）

摘　要：［**目的/意义**］玉米茎秆宽度是影响玉米抗倒伏能力的重要指标。玉米茎秆宽度测量存在人工采集过程繁琐、设备自动采集识别精度误差较大等问题，研究一种玉米茎秆宽度原位检测与高精度识别方法具有重要应用价值。［**方法**］采用ZED2i双目相机并将其固定在田间获取实时的玉米茎秆左目和右目图片，对原始图片进行数据增强，使用YOLOv8对玉米茎秆进行识别，再通过多次增加注意力机制（Coordinate Attention，CA）模块，和替换损失函数（Efficient IoU Loss，EIoU）的方法，进一步提高玉米茎秆的识别精度，然后通过对玉米茎秆的三维重建，获取识别框边界点在世界坐标系下的三维数据，通过距离公式计算出茎秆宽度。最后对改进后的YOLOv8模型与YOLOv8原模型、YOLOv7、YOLOv5、Faster RCNN、SSD进行对比，验证模型的识别准确性和识别精度。［**结果和讨论**］改进后的YOLOv8模型的查准率P、查全率R、平均精确率$mAP_{0.5}$、平均精确率$mAP_{0.5:0.95}$分别达到了96.8%、94.1%、96.6%、77.0%，玉米茎秆宽度原位检测宽度计算的线性回归决定系数R^2，均方根误差$RMSE$和平均绝对误差MAE分别为0.373、0.265和0.244 cm，可满足实际生产对玉米茎秆宽度测量精度的要求。［**结论**］本研究提出的基于改进YOLOv8模型的玉米茎秆宽度原位识别方法可以实现对玉米茎秆的原位准确识别，很好地解决了目前人工测量耗时费力和机器视觉识别精度较差的问题，为实际生产应用提供了理论依据。

关键词：YOLOv8；注意力机制；双目视觉；玉米茎宽检测；三维重建

中图分类号：TP391.4;TP181　　**文献标志码：**A　　**文章编号：**SA202309004

引用格式：左昊轩, 黄祺成, 杨佳昊, 孟繁佳, 李思恩, 李莉. 基于双目视觉和改进YOLOv8的玉米茎秆宽度原位识别方法[J]. 智慧农业(中英文), 2023, 5(3): 86-95. DOI：10.12133/j.smartag.SA202309004

ZUO Haoxuan, HUANG Qicheng, YANG Jiahao, MENG Fanjia, LI Sien, LI Li. In situ identification method of maize stalk width based on binocular vision and improved YOLOv8[J]. Smart Agriculture, 2023, 5(3): 86-95. DOI：10.12133/j.smartag.SA202309004 (in Chinese with English abstract)

1　引　言

田间作物表型的快速准确检测是监测作物生长状况的重要保障。玉米茎秆宽度是重要的表型参数，直接影响玉米的抗倒伏能力[1]。拔节期是玉米茎秆变化的重要时期，在拔节期对玉米茎秆宽度的监测尤为重要[2,3]。近年来，深度学习逐渐成为作物表型分析的重要研究手段，能提高农作物生产的管理效率。Xiang等[4]采用Mask R-CNN检测田间高粱茎秆的轮廓，计算值与真实值的相关系数为0.97。Zhou等[5,6]采用了最大类间方差（Otsu）图像阈值分割得到了玉米主茎，平均绝对误差在92.5 mm，均方根误差为25.2 mm。陈燕等[7]构建了基于深度学习架构融合ResNet和Unet的语义分割Res-Unet网络模型，分割出的玉米截面精度相较于原模型有明显提高。徐胜勇等[8]使用Mask R-CNN分割黄瓜近红外图像中的叶片和茎秆，在黄瓜三个生长时期的实际值与预测值的平均绝对误差均小于9%。张凯琪[9]使用D435i深度相机获取盆

收稿日期：2023-09-01

基金项目：国家重点研发计划支持项目（2022YFD1900801）

作者简介：左昊轩，研究方向为精细农业系统集成研究。E-mail：1226604034@qq.com

*通信作者：李 莉，副教授，博士生导师，研究方向为智慧农业系统集成和农业信息获取技术研究。E-mail：lily@cau.edu.cn

栽玉米的图像，通过RGB图像配合深度图像来进行玉米茎秆的分割，通过深度信息进行玉米茎粗的计算，实验结果的平均误差为1.36 mm。彭程等[10]提出了基于三维点云的田间香蕉吸芽表型参数信息提取方法，并针对吸芽茎秆宽度小，使用曲面拟合算法提高茎秆宽度测量精度和准确性，均方根误差为4.44 mm。

YOLOv8作为YOLO系列近期提出的版本，不仅继承了旧版本识别精度良好，识别速度较快的优点，还提升了模型本身的性能，可以被更加广泛地应用到实际生产中。袁红春和陶磊[11]提出了一种改进YOLOv8算法的商业渔船监控中鱼类的监测与识别方法。YOLOv8的检测效果对比其他主流的目标检测模型效果有明显提升。YOLOv8通过提高检测算法的精度，使得检测到的目标更准确[12, 13]。前人的目标检测都是在简单环境下完成的，在大田间复杂环境下需要对模型进行改进，完成复杂环境下的识别任务。

之前获取数据的方式多为RGB-D相机，也就是深度相机，通过结构光或ToF（Time of Fly）的物理方式获取物体的深度信息。但其有明显的局限性：测量范围窄、噪声大、视野小、易受日光干扰，因此通常在室内使用，很难应用到室外。ZED2i双目相机是一款典型的室外测距相机，由于其良好的抗强光能力、视野范围宽阔、拍摄图片清晰度高等优点，可以满足大田实际生产需求。前人的研究中对于植物表型数据计算的精度不高，识别准确度较低，计算误差不符合实际生产需求。本研究基于YOLOv8原始模型提出了一种改进的玉米茎秆宽度原位识别检测模型，可提升原始模型的识别准确率和识别精度。通过引入注意力机制，增强算法的特征提取能力，替换损失函数优化模型，保证了玉米茎秆宽度识别的精度和时效性，为玉米的茎秆宽度检测提供一种原位检测方法。

2　材料与方法

2.1 样本采集及预处理

2.1.1　样本采集

试验数据集于2023年6月8日至7月25日于甘肃省武威市中国农业大学石羊河实验站玉米表型实验田进行采集，实验田长150 m，宽50 m。玉米品种为“先玉1225”，玉米图像采集使用ZED2i双目相机，利用支架将摄像头平行于田间玉米种植行固定放置，将摄像头放置在距离地面50 cm的位置，与玉米距离40 cm，主要拍摄位置为玉米茎秆。

使用相机在田间对玉米茎秆进行拍摄，随机选取了10株玉米，使用标识物进行标记；在相机拍摄具有标识物的玉米时，使用人工测量被标识玉米茎秆的方式获取玉米茎秆的真实值。在手动测量玉米茎秆时，测量角度尽量与拍摄角度保持一致，摄像机安放示意图如图1所示。

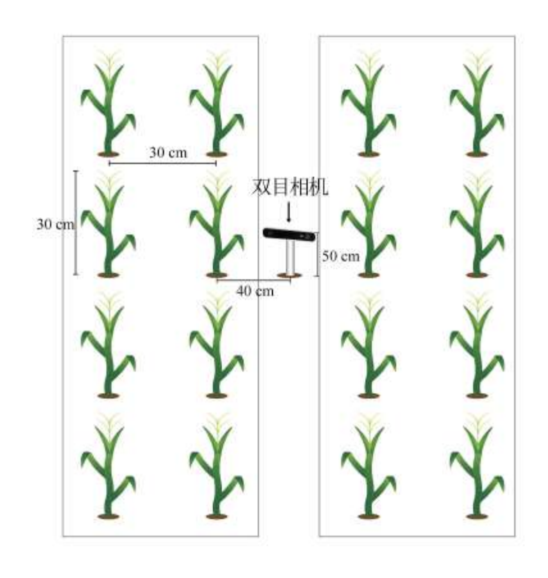


图1　田间双目相机玉米茎秆图像采集方案

Fig. 1 Binocular camera field maize stalk image acquisition scheme

使用NVIDIA Jetson TX2 NX开发板设置程序定时拍照采集，获取玉米的侧视图，共收集晴天、阴天、雨天等不同天气状况下的图像共1535幅，图像分辨率像素为2560×1920，双目相机采集图像如图2所示。



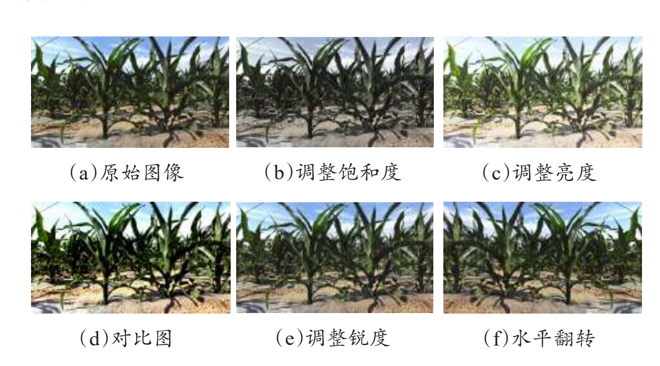
(a) 左目图片　　(b)右目图片

图2　双目相机玉米茎秆数据集图片样本

Fig. 2 Binocular camera maize stalk dataset image samplet

2.1.2 样本预处理

为了让模型能观察到目标区域的更多特征，提高模型的泛化能力，对数据集中的图像进行了数据增强。采用图像饱和度、亮度、对比度、锐度和水平翻转五种处理方式对图像进行处理，将数据集扩充至3500幅图像。增强后的数据图像样式如图3所示。



(a)原始图像 (b)调整饱和度 (c)调整亮度

(d)对比图 (e)调整锐度 (f)水平翻转

图3 数据增强后的玉米茎秆数据集图像

Fig. 3 Maize stalk dataset picture after data enhanced

将增强数据集按照8∶1∶1的比例划分为训练集、验证集和测试集。使用开源工具LabelImg对数据集进行标注，玉米种植行为东西向，双目相机放置的位置为玉米种植行南侧，水平于地面，后排存在叶片遮挡严重且土壤干扰较严重等问题，为了排除一些干扰，选择标注玉米拍摄行清晰可见完整无遮挡的玉米茎秆。使用矩形框选出玉米茎秆，同时在LabelImg软件中设置标签为stem，图像标注结果如图4所示。



图4 玉米茎秆的LabelImg标注结果

Fig.4 LabelImg labeling results of maize stalks

通过LabelImg标注后的图像保存为TXT文件。将训练集、验证集和测试集的文件名分别写入TXT文件中，将3个TXT文件存储在一个文件夹中，建立玉米茎秆宽度检测图像数据集。

2.2 识别模型构建

2.2.1 YOLOv8算法模型简介

YOLOv8属于单阶段目标检测网络[14]，YOLOv8将YOLOv5中Backbone里的C3模块更换成了梯度流更丰富的C2f模块，缩减了骨干网络中最大stage的blocks数，进一步减少参数量和计算量，实现轻量化。将YOLOv5中PAN-FPN上采样阶段中的卷积结构删除，提高YOLOv8的运算速度。YOLOv8采用的是Anchor-Free检测方式，即直接预测目标的中心点和宽高比例，而不是预测Anchor框的位置和大小。这种方式可以减少Anchor框的数量，提高检测速度和精度。

2.2.2 注意力机制

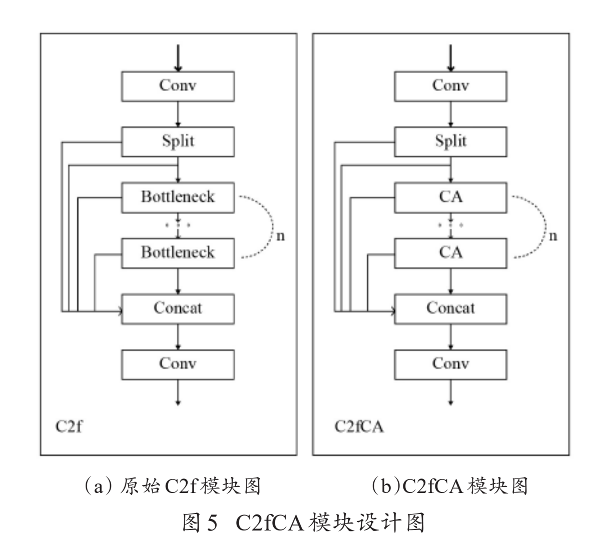
最近几年注意力机制[15]广泛使用在深度学习的各个领域，当前构建轻量级神经网络多采用挤压注意力（Squeeze-and-Excitation，SE）[16]模块，但SE模块忽略了图像的位置信息，导致图像识别会出现偏差。卷积注意力（Convolutional block attention module，CBAM）模块是一种简单而有效的前馈卷积神经网络注意模块。但CBAM仅能提取局部的位置注意力信息，缺乏长距离关系提取的能力。渠道注意力（Coordinate Attention，CA）[27]机制是一种新的高效注意力机制，不仅考虑了通道信息，还考虑了方向相关的位置信息。CA注意力机制可以在轻量级网络的基础上给下游任务带来巨大的增益，可以使注意块捕捉单方向上的长距离关系的同时保留另一个方向上的空间信息，这样位置信息就可以保存在生成的注意力图中，来关注感兴趣的区域，帮助网络更好、更准确地定位目标。

本研究网络设计中，将CA模块与原始Backbone中的C2f模块融合，利用CA注意力模块替换原始C2f模块中的Bottleneck，重新设计后的C2fCA网络模块如图5所示。

2.2.3 损失函数

在使用YOLOv8进行目标检测时，目标边界框的位置对后续的宽度计算具有重要的作用，为了缩小计算预测值与真实值的差距，改进损失函数就是一个重要的指标。

YOLOv8中box-IoU默认使用的函数为CIoU[18]（Complete IoU Loss），CIoU就是在DIoU[19]（Dis-



（a）原始C2f模块图　　（b）C2fCA模块图

图5　C2fCA模块设计图

Fig. 5 C2fCA module design diagram

tance IoU Loss）的基础上加入了宽高比，计算方法如公式（1）所示。

$$CIoU = \mathrm{IoU} - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \tag{1}$$

其中，IoU（Intersection over Union）为交并比，是测量在特定数据集中检测相应物体准确度的一个标准，它可以反映预测检测框与真实检测框的检测效果；(b，b^{gt})代表了预测框与真实框的中心点；ρ表示两个中心点之间的欧氏距离，像素；c表示的是能够同时包含预测框与真实框的最小外接矩形的对角线长度，像素；α是权重系数；v用来衡量预测框和真实框之间的宽高比的相似性，如公式（2）和（3）所示。

$$\alpha = \frac{v}{1 - \mathrm{IoU} + v} \tag{2}$$

$$v = \frac{4}{\pi^2}\left(\arctan^{-1}\frac{w^{gt}}{h^{gt}} - \arctan^{-1}\frac{w}{h}\right)^2 \tag{3}$$

其中，w^{gt}、h^{gt}分别代表真实框的宽和高，像素；w、h分别表示预测框的宽和高，像素。

最终CIoU Loss按照公式（4）计算。

$$L_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{4}$$

其中，L_{CIoU}为CIoU损失。CIoU仅能反映预测边界框与真实边界框长宽比的差异，并不是长宽分别与其置信度的真实差异，预测边界框与真实边界框长宽比描述的是相对值，可能会导致CIoU以不合理的方式优化相似性，并未考虑难易样本的平衡问题。针对这一问题，有学者提出了EIoU（Efficient IoU Loss）。EIoU[20]是在CIoU的惩罚项基础上将预测框和真实框的纵横比的影响因子拆开，分别计算预测框和真实框的长和宽，来解决CIoU存在的问题，其按照公式（5）计算。

$$L_{\mathrm{EIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \tag{5}$$

其中，L_{EIoU}为EIoU损失；c_w和c_h是预测框和真实框最小外接矩形的宽度和高度，像素。

EIoU将纵横比的损失项拆分成预测的宽高分别与最小外接框宽高的差值，加速了预测框的收敛、提高了预测框的回归精度。

2.2.4　改进后的网络

本研究以YOLOv8为基础模型加以改进，以适应田间玉米在复杂环境背景下的茎秆原位识别检测，基于以上叙述及试验论证，针对YOLOv8的改进如下：将Backbone的C2f模块中的Bottleneck前增加CA注意力，可以使网络更精准地定位目标，并提高网络的效率。然后更换损失函数，进一步提高收敛速度从而优化网络提高网络的识别精度。改进后的网络结构如图6所示。

2.3 玉米茎秆宽度计算

2.3.1　双目相机标定

双目立体视觉是用两个相机在不同角度同时拍摄物体，通过两幅图像视觉差恢复物体的三维几何形状。双目相机成像原理如图7所示。

对双目相机进行标定，双目相机标定采用棋盘格标定，通过双目相机对棋盘格不同方向的拍摄获取图像。共采集24张不同角度的图片，使用OpenCV-Python自动双目标定可获取双目相机的内参，外参和畸变指数。通过标定结果对原始图像进行矫正，矫正后的两张图像位于同一平面，可为后续三维重建提供更加精确的数据。

2.3.2　三维重建计算玉米茎秆宽度

使用1.2节改进的YOLOv8在识别出玉米茎秆后，继续识别出图像中识别框左上角点和右下角点的坐标信息，通过双目视觉对识别出的玉米茎秆进行三维重建，对左右两目图像识别框坐标信息、边界框宽度和高度以及置信度的比较匹配，完成双目图像中的识别框匹配。玉米茎秆宽度估算检测算法流程图如图8所示。

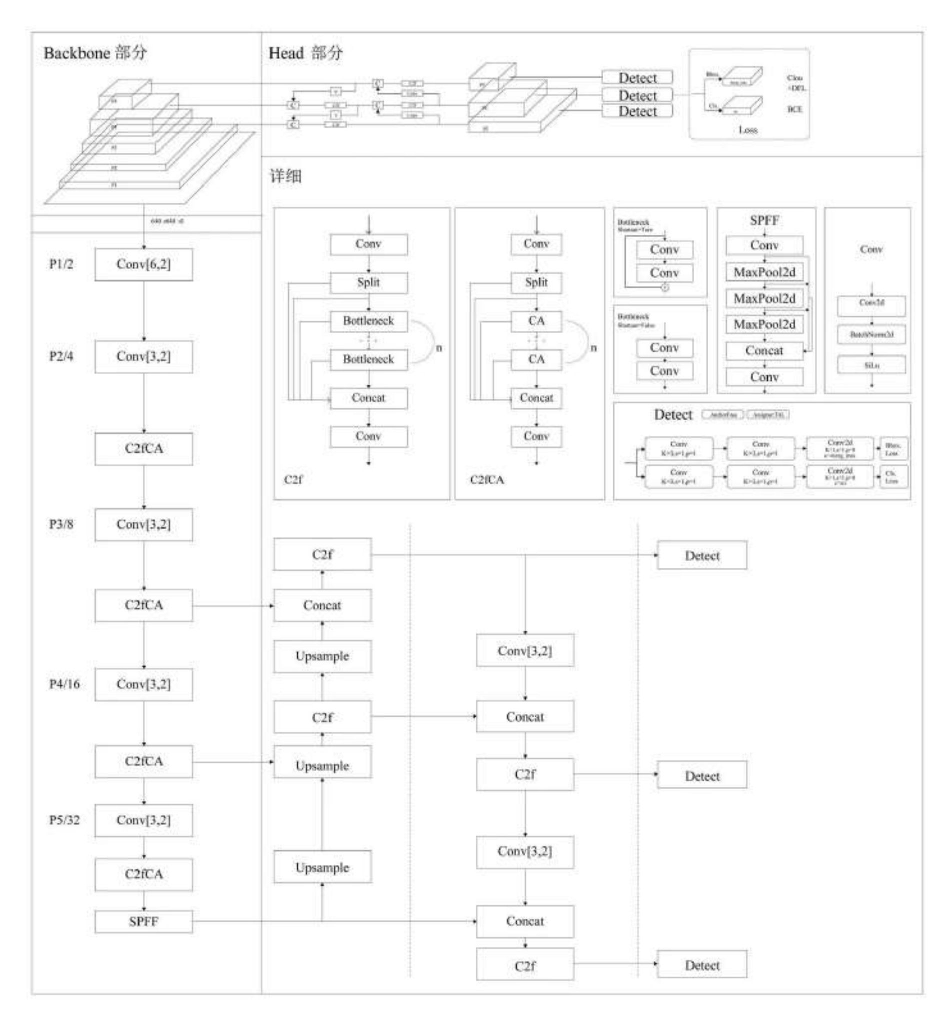


图6 改进后的YOLOv8网络结构图

Fig. 6 Structure of the improved YOLOv8 network

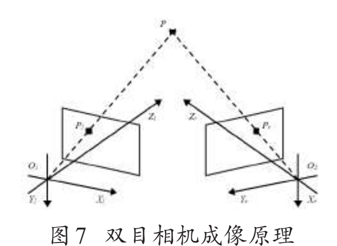


图7 双目相机成像原理

Fig.7 Binocular camera imaging principle

由图8所示，首先判断两张图像中识别框的检测数量是否相等，若不相等，则重新输入双目图像；若相等，则继续判断左右目图像的坐标信息、边界框的宽度和高度，判断差值是否小于给定阈值a（Threshold a，T_a）。若大于给定T_a，则重新输入图像；若小于给定T_a，则通过判断图像的识别框的置信度是否小于给定阈值b（Threshold b，T_b）。若大于给定T_b，则重新输入图像；若小于给定T_b，则表明该识别框为左右图像中识别的同一株玉米。满足以上条件则完成双目图像中的对应点匹配。在双目图像三维重建后，获得识别框左上角与右上角在世界坐标系下的三维坐标（A_x，A_y，A_z）和（B_x，B_y，B_z），两点的距离就是玉米茎秆的宽度W，计算方法如公式（6）所示。

$$W = \sqrt{\left(A_x - B_x\right)^2 + \left(A_y - B_y\right)^2 + \left(A_z - B_z\right)^2} \tag{6}$$

其中，（A_x，A_y，A_z）是识别框左上角在世界坐标系下的坐标；（B_x，B_y，B_z）是识别框右上角在世界坐标系下的坐标。

3 试验与结果分析

3.1 试验环境及参数设置

该试验在Windows10系统上运行，配置为搭载Intel Core I5-10400F CPU，主频2.9 GHz，最高睿频

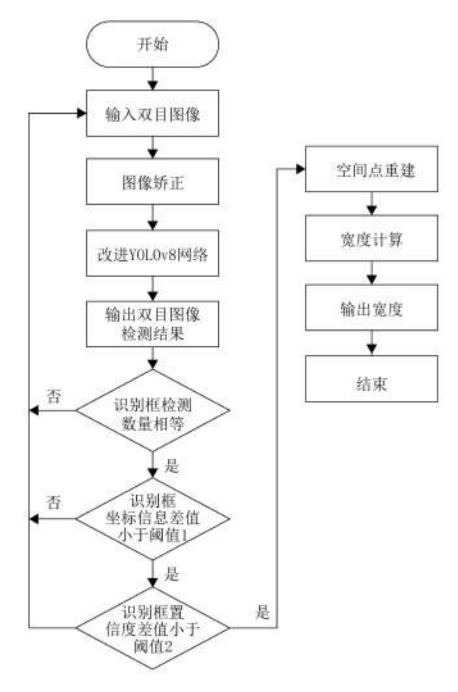


图8　玉米茎秆宽度估测算法流程图

Fig.8 Flow chart of maize stalk width estimation algorithm

4.3 GHz，32 G机带RAM，Nvidia Geforce RTX 3070显卡，统一计算设备架构（CUDA）版本为11.3，GPU加速CUDNN版本为8.2.0，编程语言为Python3.8，深度学习框架为Pytorch1.10.0。

初始学习率设为0.01，动量初始值0.937，权重衰减系数为0.0005，图像输入像素尺寸为2560×1920，batch-size为35，训练轮次设为500个epochs，IoU阈值0.5。

3.2 试验评价指标

本研究主要采用查准率（Precision，P），召回率（Recall，R），P-R曲线在二维坐标中围起来的面积大小，被称作平均准确率（Average Precision，AP），多对象的平均精确率为mAP，因为研究对象只有一个，所以文中均用mAP作为比较指标，如公式（7）所示。

$$mAP = \int_0^1 PdR \tag{7}$$

其中，P是查准率，dR是对召回率进行积分。平均精度在IoU为0.5时（mean Average Precision，$mAP_{0.5}$），以及平均精度在IoU从0.5到0.95，步长0.05（mean Average Precision，$mAP_{0.5:0.95}$）作为评价标准，将提出的改进模型与其他模型进行了对比，其中查准率与召回率按照公式（8）和（9）计算。

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

其中，TP（True Positives）表示实际玉米茎秆被判别为玉米茎秆的样本数量，个；FP（False Positives）表示实际为非玉米茎秆被判别为玉米茎秆的样本数量，个；FN（False Negatives）表示实际为玉米茎秆被判别为非玉米茎秆的样本数量，个；TN（True Negatives）表示实际为非玉米茎秆被判别为非玉米茎秆的样本数量，个。

3.3 性能结果分析

为了检验对YOLOv8算法改进的效果，通过进行消融实验来评估这些改进对整体性能的影响。首先对注意力机制进行评估，通过对比多项常见注意力机制模块，如SE注意力模块、CBAM注意力模块以及本研究的C2fCA注意力模块，按照上述的改进方法以及训练策略，在训练完成后，在同一个测试集上进行推理验证，实际验证结果表明，在本次试验中C2fCA注意力模块的识别效果最优，验证集上的识别准确率如图9所示。

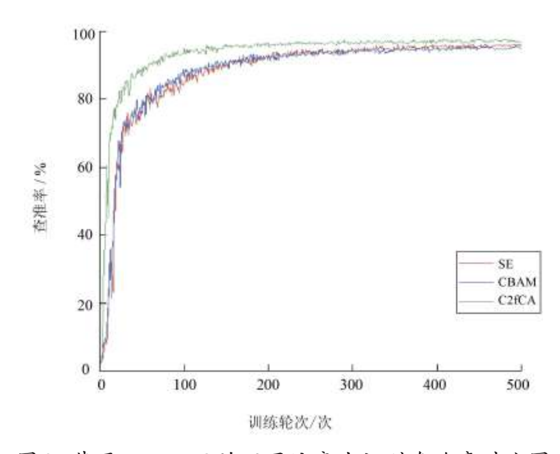


图9　基于YOLOv8的不同注意力识别准确率对比图

Fig.9 Comparison chart of different attention recognition accuracy rates based on YOLOv8

在C2fCA的模型基础上，更换YOLOv8的损失函数，YOLOv8中box-IoU默认使用的函数为CIoU，将损失函数中的CIoU替换为EIoU，分别对比了EIoU与CIoU的各项评估指标，在准确率和召回

率上二者并无明显区别，但在预测框与真实框之间的误差回归中，EIoU的收敛速度和精度要比CIoU表现优秀，二者对比图如图10所示。

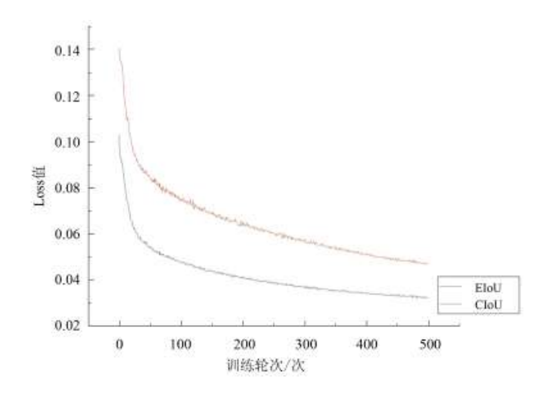


图10 基于YOLOv8的EIoU与CIoU回归Loss对比图

Fig. 10 Comparison chart of EIoU and CIoU regression loss based on YOLOv8

不同改进的消融实验的结果相较于原始YOLOv8各项参数指标如表1所示。

由表1可知，经过多项消融实验后，改进后的YOLOv8算法对比两项改进均不添加的原始YOLOv8算法，P提升了2.1%，R提升了1.6%，$mAP_{0.5}$提升了2.2%，$mAP_{0.5:0.95}$提升了14.4%。实验结果表明，增加CA注意力机制，更换损失函数可以提高整个检测网络的识别精度，同时使用CA和EIoU后，模型的画面每秒传输帧数（Frame Per Second，FPS）相比单独使用CA和EIoU更低，是因为CA或EIoU的引入会增加模型的复杂性，导致处理每一帧的时间增加，在牺牲一定FPS的情况下，提高模型的精度。FPS大于等于24即可满足实时检测的需求，本研究提出的改进YOLOv8算法在提升玉米茎秆的识别率的同时，又保证了检测速度。

基于YOLOv8的玉米茎秆宽度测量结果如图11所示。训练结果可以在边界框中显示茎秆宽度数据。其中，通过在田间定位标记植株并标记图片的方式来选取了10株玉米，通过手动测量和算法计算来分析茎秆宽度预测与实际测量的误差。表2比较了茎秆宽度预测与实际测量数据分析的决定系数（R-Square，R^2），均方根误差（Root Mean Square Error，RSME）和绝对平均误差（Mean Absolute Error，MAE）的分析结果。由表2可知，通过YOLOv8估计得到的玉米茎秆宽度的值R^2为0.373，RMSE为0.265 cm，MAE为0.244 cm。R^2可以衡量模型对数据的拟合程度，尽管该模型的R^2为0.373，但是模型依旧拥有良好的预测结果，该结果可以应用在实际生产，该误差在实际大田原位检测茎秆宽度中是可接受的误差范围。



图11 基于YOLOv8的玉米茎秆宽度测量结果

Fig. 11 Measurement results of maize stalk width based on YOLOv8

玉米茎秆宽度原位检测由于玉米在实际种植过程中并不都是直立生长，在图像获取过程中会存在轻微的歪斜，获取图像时会存在误差；人工对茎秆宽度进行测量时，由于原位测量存在难度，在测量时也会存在误差，这些都属于数据的不确定性。在识别过程中由于玉米前后茎秆的重叠会被误判为一个茎秆，会使识别效果出现误差。

3.4 不同模型的对比试验分析

为进一步验证本研究提出的YOLOv8改进算法

表1 YOLOv8的不同改进消融实验结果

Table 1 Results of different improved ablation tests of YOLOv8

模型			P/%	R/%	$mAP_{0.5}$/%	$mAP_{0.5:0.95}$/%	FPS
	CA	EIoU	——	——	——	——	——
YOLOv8	×	×	94.7	92.5	94.4	62.6	69.0
	√	×	96.2	93.5	96.1	70.5	57.0
	×	√	95.3	92.5	95.9	68.8	58.0
	√	√	96.8	94.1	96.6	77	56.0

表2　基于YOLOv8的玉米茎秆宽度测量结果分析

Table 2　Analysis of maize stalk width measurement results based on YOLOv8

玉米植株编号	真值/cm	检测值/cm	偏差/cm
1	2.34	2.25	0.09
2	2.82	2.36	0.46
3	2.47	2.30	0.17
4	2.32	2.29	0.03
5	2.56	2.28	0.28
6	2.15	2.32	−0.17
7	2.23	2.30	−0.07
8	1.93	2.25	−0.32
9	2.54	2.33	0.21
10	2.02	2.29	−0.27
R^2		0.373	
RMSE/cm		0.265	
MAE/cm		0.244	

模型的性能，对比其他当前主流的卷积神经网络模型，试验采用YOLOv7、YOLOv5、更快的区域卷积神经网络（Faster Region Convolutional Neural Networks，Faster R-CNN）和单次多盒检测器（Single Shot MultiBox Detector，SSD）在上述同一参数设置条件下，在自建玉米茎秆识别数据集的训练集和测试集上进行训练和测试，最终得到的性能测试结果如表3所示。

表3　玉米茎秆检测的不同模型对比结果

Table 3　Comparison results of different models for maize stalk detection

算法	P/%	R/%	$mAP_{0.5}$/%	$mAP_{0.5:0.95}$/%
YOLOv8	96.8	94.1	96.6	77.0
YOLOv7	95.5	92.8	95.6	65.4
YOLOv5	95.0	92.0	95.4	61.2
Faster R-CNN	65.7	53.8	50.4	39.4
SSD	76.2	70.3	75.7	56.9

由表3可知，本研究提出的YOLOv8改进模型在玉米茎秆宽度检测验证数据集上对比其他几种主流卷积神经网络模型，识别准确率和召回率可达到96.8%和94.1%，$mAP_{0.5}$达到了96.6%，$mAP_{0.5:0.95}$达到了77.0%，对比YOLOv7分别提升了1.3%、1.3%、1.0%、11.6%，对比YOLOv5分别提升了1.8%、2.1%、1.2%、15.8%，对比Faster R-CNN分别提升了31.1%、40.3%、46.2%、37.6%，对比SSD分别提升了20.6%、23.8%、20.9%、20.1%。在算法性能上有着更好的表现效果，说明改进后的YOLOv8模型可以更好地定位到目标区域，排除背景复杂的干扰，也能够提升识别的准确率和识别速度，可以充分利用图像信息。

4　结　论

本研究建立了玉米茎秆宽度原位检测数据集，在YOLOv8模型的基础上提出了一种融合CA注意力机制，通过引导网络加强对目标区域的检测，减少农业生产中复杂背景的影响，降低背景干扰对识别效果的影响，替换YOLOv8交叉熵函数中的CIoU为EIoU，加强预测框与真实框之间的收敛速度，提高玉米茎秆识别检测精度。利用改进的YOLOv8算法进行训练后对玉米茎秆进行识别，最终得出识别准确率、召回率、$mAP_{0.5}$和$mAP_{0.5:0.95}$分别为96.8%、94.1%、96.6%和77.0%，相较于原YOLOv8，P提升了2.1%，R提升了1.6%，$mAP_{0.5}$提升了2.2%，$mAP_{0.5:0.95}$提升了14.4%。通过双目图像深度信息估计得到的玉米茎秆宽度的值R^2为0.343，RMSE为0.333 cm，MAE为0.270 cm，此计算精度能够充分满足农业生产的要求，可以达到玉米茎秆原位检测的要求。本研究提出的基于改进的YOLOv8的玉米茎秆宽度原位识别方法可以满足实际生产需要。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] WANG J Z, LI X H, ZHOU Y H, et al. Banana pseudostem width detection based on kinect V2 depth sensor[J]. Computational intelligence and neuroscience, 2022, 2022: ID 3083647.

[2] DU J, ZHANG Y, LU X, et al. A deep learning-integrated phenotyping pipeline for vascular bundle phenotypes and its application in evaluating sap flow in the maize stem[J]. The crop journal, 2022, 10(5): 1424-1434.

[3] 胡松涛, 翟瑞芳, 王应华, 等. 基于多源数据的马铃薯植株表型参数提取[J]. 智慧农业(中英文), 2023, 5(1): 132-145.
HU S T, ZHAI R F, WANG Y H, et al. Extraction of potato plant phenotypic parameters based on multi-source data[J]. Smart agriculture, 2023, 5(1): 132-145.

[4] XIANG L, TANG L, GAI J, et al. Measuring stem diameter of sorghum plants in the field using a high-throughput stereo vision system[J]. Transactions of the ASABE, 2021, 64(6): 1999-2010.

[5] ZHOU J, CUI M R, WU Y S, et al. Maize (*Zea mays* L.) stem target region extraction and stem diameter measurement based on an internal gradient algorithm in field conditions[J]. Agronomy, 2023, 13(5): ID 1185.

[6] ZHOU J, WU Y S, CHEN J A, et al. Maize stem contour extraction and diameter measurement based on adaptive threshold segmentation in field conditions[J]. Agriculture, 2023, 13(3): ID 678.

[7] 陈燕, 李想, 曹勉, 等. 基于语义分割与实例分割的玉米茎秆截面参数测量方法[J]. 农业机械学报, 2023, 54(6): 214-222.
CHEN Y, LI X, CAO M, et al. Measurement of maize stem cross section parameters based on semantic segmentation and instance segmentation[J]. Transactions of the Chinese society for agricultural machinery, 2023, 54(6): 214-222.

[8] 徐胜勇, 李磊, 童辉, 等. 基于RGB-D相机的黄瓜苗3D表型高通量测量系统研究[J]. 农业机械学报, 2023, 54(7): 204-213, 281.
XU S Y, LI L, TONG H, et al. High-throughput measurement system for 3D phenotype of cucumber seedlings using RGB-D camera[J]. Transactions of the Chinese society for agricultural machinery, 2023, 54(7): 204-213, 281.

[9] 张凯琪. 基于深度学习的盆栽玉米茎粗自动测量平台的研究[D]. 武汉: 华中农业大学, 2022.
ZHANG K Q. Research on automatic measurement platform of potted maize stem diameter based on deep learning[D]. Wuhan: Huazhong Agricultural University, 2022.

[10] 彭程, 苗艳龙, 汪刘洋, 等. 基于三维点云的田间香蕉吸芽形态参数获取[J]. 农业工程学报, 2022, 38(S1): 193-200.
PENG C, MIAO Y L, WANG L Y, et al. Morphological parameters extraction of banana sucker in the field based on three-dimensional point cloud[J]. Transactions of the Chinese society of agricultural engineering, 2022, 38(S1): 193-200.

[11] 袁红春, 陶磊. 基于改进的YOLOv8商业渔船电子监控数据中鱼类的检测与识别[J]. 大连海洋大学学报, 2023, 38(3): 533-542.
YUAN H C, TAO L. Detection and identification of fish in electronic monitoring data of commercial fishing vessels based on improved YOLOv8[J]. Journal of Dalian Ocean university, 2023, 38(3): 533-542.

[12] LI Y T, FAN Q S, HUANG H S, et al. A modified YOLOv8 detection network for UAV aerial image recognition[J]. Drones, 2023, 7(5): ID 304.

[13] MARTINEZ-CARRANZA J, HERNÁNDEZ-FARÍAS D I, ROJAS-PEREZ L O, et al. Language meets YOLOv8 for metric monocular SLAM[J]. Journal of real-time image processing, 2023, 20(4): 1-10.

[14] LOU H T, DUAN X H, GUO J M, et al. DC-YOLOv8: Small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): ID 2323.

[15] PRINZMETAL W, HA R, KHANI A. The mechanisms of involuntary attention[J]. Journal of experimental psychology human perception and performance, 2010, 36(2): 255-267.

[16] ZHONG X, GONG O B, HUANG W X, et al. Squeeze-and-excitation wide residual networks in image classification[C]// 2019 IEEE International Conference on Image Processing (ICIP). Piscataway, New Jersey, USA: IEEE, 2019: 395-399.

[17] LI X Z, WU B Y, ZHU X, et al. Consecutively missing seismic data interpolation based on coordinate attention unet[J]. IEEE geoscience and remote sensing letters, 2022, 19: 1-5.

[18] XUE J L, CHENG F, LI Y Q, et al. Detection of farmland obstacles based on an improved YOLOv5s algorithm by using CIoU and anchor box scale clustering[J]. Sensors, 2022, 22(5): ID 1790.

[19] YINAN W, YUN Z, JIA G, et al. YOLOv5 detection algorithm of steel defects based on introducing light convolution network and DIOU function[C]// 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS). Piscataway, New Jersey, USA: IEEE, 2023: 118-122.

[20] LI Z Y, JIANG X Q, SHUAI L Y, et al. A real-time detection algorithm for sweet cherry fruit maturity based on YOLOX in the natural environment[J]. Agronomy, 2022, 12(10): ID 2482.

In Situ Identification Method of Maize Stalk Width Based on Binocular Vision and Improved YOLOv8

ZUO Haoxuan[1], HUANG Qicheng[1], YANG Jiahao[2], MENG Fanjia[2], LI Sien[3], LI Li[1*]

(1. *Key Laboratory of Agricultural Information Acquisition Technology*, *Ministry of Agriculture and Rural Affairs*, *China Agricultural University*, *Beijing* 100083, *China*; 2. *Key Laboratory of Smart Agriculture System Integration*, *Ministry of Education*, *China Agricultural University*, *Beijing* 100083, *China*; 3. *College of Water Resources and Civil Engineering*, *China Agricultural University*, *Beijing* 100083, *China*)

Abstract:
[Objective] The width of maize stalks is an important indicator affecting the lodging resistance of maize. The measurement of maize stalk width has many problems, such as cumbersome manual collection process and large errors in the accuracy of automatic equipment collection and recognition, and it is of great application value to study a method for in-situ detection and high-precision identification of maize stalk width.
[Methods] The ZED2i binocular camera was used and fixed in the field to obtain real-time pictures from the left and right sides of maize stalks together. The picture acquisition system was based on the NVIDIA Jetson TX2 NX development board, which could achieve timed shooting of both sides view of the maize by setting up the program. A total of maize original images were collected and a dataset was established. In order to observe more features in the target area from the image and provide assistance to improve model training generalization ability, the original images were processed by five processing methods: image saturation, brightness, contrast, sharpness and horizontal flipping, and the dataset was expanded to 3500 images. YOLOv8 was used as the original model for identifying maize stalks from a complex background. The coordinate attention (CA) attention mechanism can bring huge gains to downstream tasks on the basis of lightweight networks, so that the attention block can capture long-distance relationships in one direction while retaining spatial information in the other direction, so that the position information can be saved in the generated attention map to focus on the area of interest and help the network locate the target better and more accurately. By adding the CA module multiple times, the CA module was fused with the C2f module in the original Backbone, and the Bottleneck in the original C2f module was replaced by the CA module, and the C2fCA network module was redesigned. Replacing the loss function Efficient IoU Loss(EIoU) splits the loss term of the aspect ratio into the difference between the predicted width and height and the width and height of the minimum outer frame, which accelerated the convergence of the prediction box, improved the regression accuracy of the prediction box, and further improved the recognition accuracy of maize stalks. The binocular camera was then calibrated so that the left and right cameras were on the same three-dimensional plane. Then the three-dimensional reconstruction of maize stalks, and the matching of left and right cameras recognition frames was realized through the algorithm, first determine whether the detection number of recognition frames in the two images was equal, if not, re-enter the binocular image. If they were equal, continue to judge the coordinate information of the left and right images, the width and height of the bounding box, and determine whether the difference was less than the given T_a. If greater than the given T_a, the image was re-imported; If it was less than the given T_a, the confidence level of the recognition frame of the image was determined whether it was less than the given T_b. If greater than the given T_b, the image is re-imported; If it is less than the given T_b, it indicates that the recognition frame is the same maize identified in the left and right images. If the above conditions were met, the corresponding point matching in the binocular image was completed. After the three-dimensional reconstruction of the binocular image, the three-dimensional coordinates (A_x, A_y, A_z) and (B_x, B_y, B_z) in the upper left and upper right corners of the recognition box under the world coordinate system were obtained, and the distance between the two points was the width of the maize stalk. Finally, a comparative analysis was conducted among the improved YOLOv8 model, the original YOLOv8 model, faster region convolutional neural networks (Faster R-CNN), and single shot multiBox detector (SSD)to verify the recognition accuracy and recognition accuracy of the model.
[Results and Discussions] The precision rate (P)、recall rate (R)、average accuracy $mAP_{0.5}$、average accuracy $mAP_{0.5:0.95}$ of the improved YOLOv8 model reached 96.8%、94.1%、96.6% and 77.0%. Compared with YOLOv7, increased by 1.3%、1.3%、1.0% and 11.6%, compared with YOLOv5, increased by 1.8%、2.1%、1.2% and 15.8%, compared with Faster R-CNN, increased by 31.1%、40.3%、46.2%、and 37.6%, and compared with SSD, increased by 20.6%、23.8%、20.9% and 20.1%, respectively. Respectively, and the linear regression coefficient of determination R^2, root mean square error RMSE and mean absolute error MAE were 0.373, 0.265 cm and 0.244 cm, respectively. The method proposed in the research can meet the requirements of actual production for the measurement accuracy of maize stalk width.
[Conclusions] In this study, the in-situ recognition method of maize stalk width based on the improved YOLOv8 model can realize the accurate in-situ identification of maize stalks, which solves the problems of time-consuming and laborious manual measurement and poor machine vision recognition accuracy, and provides a theoretical basis for practical production applications.

Key words: YOLOv8; attention mechanism; binocular vision; maize stalk width detection; three-dimensional reconstruction

Foundation Items: National Key Research and Development Program of China (2022YFD1900801)